Chapter 2

Machine Translation in Practice

2.1 Introduction

At the time of writing, the use of MT — or indeed, any sort of computerised tool for translation support — is completely unknown to the vast majority of individuals and organizations in the world, even those involved in the so called ‘language industries’, like translators, terminologists, technical writers, etc.

Given this, one of the first things a reader is likely to want to know about MT is what it might be like to work with an MT system and how it fits in with the day-to-day business of translation. The purpose of the present chapter is to provide just such information — a view of MT at the user level, and from the outside. In later chapters we shall in effect lift off the covers of an MT system and take a look at what goes on inside. For the moment, however, the central components of an MT system are treated as a black box.

We introduce the business of MT in terms of a scenario describing the usage of MT inside a fairly large multinational corporation. The scenario is not based exactly on any one existing corporation. Our description is somewhat idealised in that we assume methods of working which are only just starting to come into use. However, there is nothing idly futuristic in our description: it is based on a consensus view of commercial MT experts and envisages tools which we know to be either already available or in an advanced state of development in Europe or elsewhere. The commercialisation of MT is not awaiting a ‘miracle breakthrough’ in the science of MT; it is not necessary, nor do we expect it to occur. What will happen over the next ten years are progressive improvements in functionality and performance which, taken in conjunction with the continuously falling costs of basic computing power, will ensure that MT becomes more and more cost effective. In short, we have no doubt that in general outline, if not in every detail, we are sketching the professional life of the machine translator in the 90s, and of most translators in the early part of the next century.

2.2 The Scenario

Let us suppose that you are a native English speaker engaged as a professional German-English translator in the Language Centre for a multinational manufacturing company. One of the products this company supplies is computer products. In this organization the Language Centre is principally responsible for the translation of documents created within the company into a variety of European and Oriental languages. The Language Centre is also charged with exercising control over the content and presentation of company documentation in general. To this end, it attempts to specify standards for the final appearance of documents in distributed form, including style, terminology, and content in general. The overall policy is enshrined in the form of a corporate *Document Design and Content Guide* which the Centre periodically updates and revises.

The material for which MT is to be used consists of technical documentation such as User and Repair manuals for software and hardware products manufactured or sourced by the company. Some classes of highly routine internal business correspondence are also submitted for MT. Legal and marketing material, and much external business correspondence, is normally translated by hand, although some translators in the organization prefer to use MT here as well.

All material for translation is available in electronic form on a computer network which supports the company's documentation system. Although most documents will be printed out at some point as standard paper User Manuals and so forth, the system also supports the preparation of multi-media hypertext documents. These are documents which exist primarily in electronic form with a sophisticated cross-reference system; they contain both text and pictures (and perhaps speech and other sounds). These documents are usually distributed to their final users as CD-ROMs, although they can be distributed in other electronic forms, including electronic mail. Printed versions of these documents can also be made.

Everyone in the language department has a workstation — an individual computer. These are linked together by the network. The documentation system which runs on this network allows users to create and modify documents by typing in text; in other words, it provides very sophisticated word processing facilities. It also provides sophisticated means for storing and retrieving electronic documents, and for passing them around the network inside the company or via external networks to external organizations. As is usual with current computer systems, everything is done with the help of a friendly interface based on windows, icons and menus, selections being made with a mouse.

The MT system which you use is called ETRANS and forms part of the overall documentation system. (ETRANS is just a name we have invented for a prototypical MT system.) Parts of an electronic document on the system can be sent to the MT system in the same way that they can be sent to a printer or to another device or facility on the network. ETRANS is simultaneously available from any workstation and, for each person using it, behaves as if it is his or her own personal MT system.

Earlier this morning, one of the technical authors had completed (two days after the deadline) a User Manual for a printer the company is about to launch. The text is in German. Although this author works in a building 50 kilometres away, the network ensures that the document is fully accessible from your workstation. What follows is a fragment of the text which you are viewing in a window on the workstation screen and which you are going to translate:

German Source Text

Druckdichte Einstellung
Die gedruckte Seite sollte von exzellenter Qualität sein. Es gibt aber eine Reihe von Umweltfaktoren, wie hohe Temperatur und Feuchtigkeit, die Variationen in der Druckdichte verursachen können.
Falls die Testseite zu hell oder zu dunkel aussieht, verstellen Sie die Druckdichte am Einstellknopf an der linken Seite des Druckers (Figur 2-25).
Einstellung der Druckdichte:

- Drehen Sie den Knopf ein oder zwei Positionen in Richtung des dunklen Indikators.
- Schalten Sie den Drucker für einen Moment aus und dann wieder ein, so daβ die Testseite gedruckt wird.
- Wiederholen Sie die beiden vorherigen Schritte solange, bis Sie grau auf dem Blatthintergrund sehen, ähnlich wie bei leicht unsauberen Kopien eines Photokopierers.
- Drehen Sie den Knopf eine Position zurück.

Jetzt können Sie den Drucker an den Computer anschliessen.
Falls Sie den Drucker an einen Macintosh Computer anschliessen, fahren Sie mit den Instruktionen im Kapitel 3 fort. Falls Sie einen anderen Computer benutzen, fahren Sie fort mit Kapitel 4.

As with all the technical documents submitted to ETRANS, all the sentences are relatively short and rather plain. Indeed, it was written in accordance with the Language Centre document specification and with MT very much in mind. There are no obvious idioms or complicated linguistic constructions. Many or all of the technical terms relating to printers (e.g. *Druckdichte* 'print density') are in regular use in the company and are stored and defined in paper or electronic dictionaries available to the company's technical authors and translators.

To start up ETRANS, you click on the icon bearing an ETRANS logo, and this pops up a

menu giving various translation options. ETRANS handles six languages: English, German, French, Italian, Spanish and Japanese. The printer document needs to be translated into English, so you select English as the target language option. Another menu shows the source language to be used. In this case, there is no need to select German because ETRANS has already had a very quick look at your printer document and decided, given rather superficial criteria such as the presence of umlauts and other characteristics of German orthography, that it is probably German text. If ETRANS had guessed wrongly — as it sometimes does — then you could select the correct source language from the menu yourself. By clicking on an additional menu of ETRANS options, you start it translating in batch or full-text mode; that is, the whole text will be translated automatically without any intervention on your part. The translation starts appearing in a separate screen window more or less immediately. However, since the full source text is quite long, it will take some time to translate it in its entirety. Rather than sit around, you decide to continue with the revision of another translation in another window. You will look at the output as soon as it has finished translating the first chapter.

The output of ETRANS can be found on page 23. The quality of this raw output is pretty much as you expect from ETRANS. Most sentences are more or less intelligible even if you don't go back to the German source. (Sometimes some sentences may be completely unintelligible.) The translation is relatively accurate in the sense that it is not misleading — it doesn't lead you to think that the source text says one thing when it really says something quite the opposite. However, the translation is very far from being a good specimen of English. For one thing, ETRANS clearly had difficulties with choosing the correct translation of the German word *ein* which has three possible English equivalents: *a*/*an*, *on* and *one*.

(1) a. Turn the button an or two positions in direction of the dark indicator.
b. Switch off the printer for a moment and then again a, so that the test page is printed.

Apart from these details, it has also made quite a mess of a whole phrase:

(2) . . . , similarly like at easily unclean copies of a photocopier.

In order to post-edit such phrases it will be necessary to refer back to the German source text.

MT Output

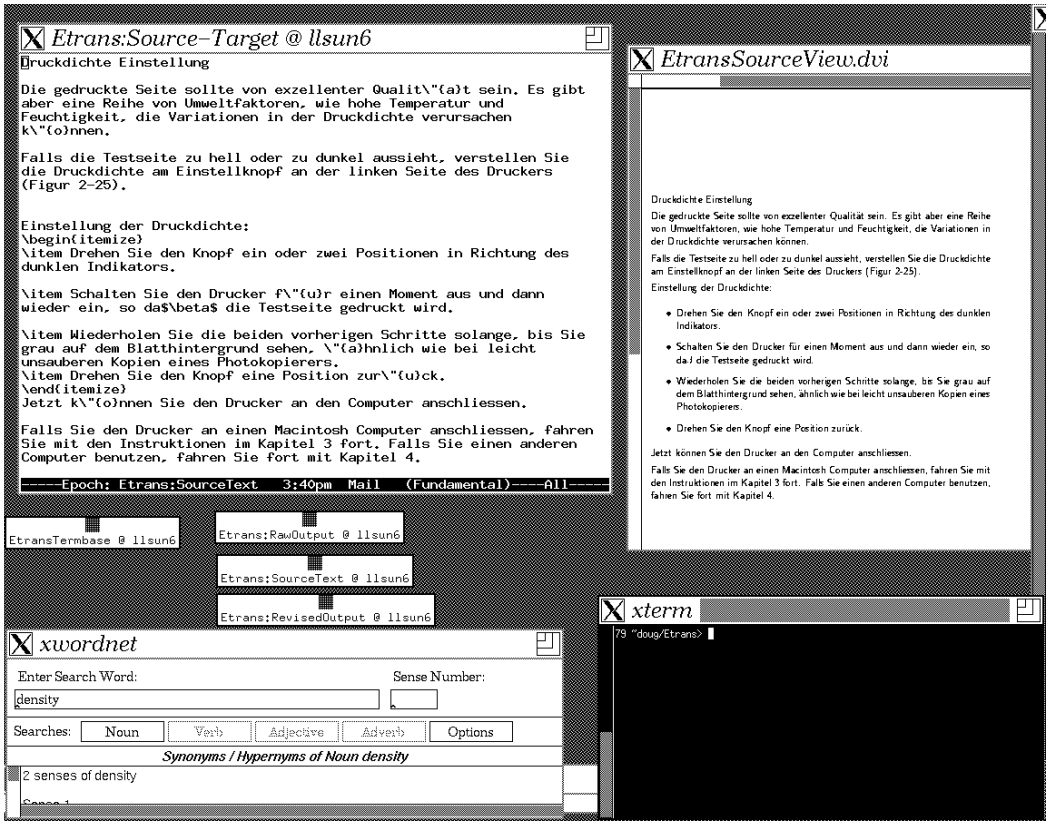
Print density adjustment
The printed page should be from excellent quality. There is however a series of environmental factors, how high temperature and humidity, can cause the variations in the print density.
If the test page looks too light or too darkly, adjust the print density at the tuner at the left page of the printer (figure 2-25).
Adjustment of the print density:

- Turn the button an or two positions in direction of the dark indicator.
- Switch off the printer for a moment and then again a, so that the test page is printed.
- Repeat the two previous steps as long as, until you see Gray on the background of the page, similarly like at easily unclean copies of a photocopier.
- Turn back the button a position.

Now you can connect the printer to the computer.
If you connect the printer to a Macintosh computers, continue with the instructions in the chapter 3. If you use an other computer, continue with chapters 4.

Leaving ETRANS to continue translating later chapters of the document, you start post-editing the first chapter by opening up a post-edit window, which interleaves a copy of the raw ETRANS output with the corresponding source sentences (e.g. so that each source sentence appears next to its proposed translation). Your workstation screen probably now looks something like the Figure on page 24.

Icons and menus give access to large scale on-line multilingual dictionaries — either the ones used by the ETRANS itself or others specifically intended for human users. You post-edit the raw MT using the range of word-processing functions provided by the document processing system. Using search facilities, you skip through the document looking for all instances of *a*, *an* or *one*, since you know that these are often wrong and may need replacement. (Discussions are in progress with the supplier of ETRANS who has promised to look into this problem and make improvements.) After two or three other global searches for known problem areas, you start to go through the document making corrections sentence by sentence. The result of this is automatically separated from the source text, and can be displayed in yet another window. Page 26 shows what your workstation screen might now look like.


Etrans:Source-Target @ llsun6
Druckdichte Einstellung
Die gedruckte Seite sollte von exzellenter Qualit\"{a}t sein. Es gibt aber eine Reihe von Umweltfaktoren, wie hohe Temperatur und Feuchtigkeit, die Variationen in der Druckdichte verursachen k\"{o}nnen.
Falls die Testseite zu hell oder zu dunkel aussieht, verstellen Sie die Druckdichte am Einstellknopf an der linken Seite des Druckers (Figur 2-25).
Einstellung der Druckdichte:
\begin{itemize}
\item Drehen Sie den Knopf ein oder zwei Positionen in Richtung des dunklen Indikators.
\item Schalten Sie den Drucker f\"{u}r einen Moment aus und dann wieder ein, so daβ die Testseite gedruckt wird.
\item Wiederholen Sie die beiden vorherigen Schritte solange, bis Sie grau auf dem Blatthintergrund sehen, \"{a}hnlich wie bei leicht unsauberen Kopien eines Photokopierers.
\item Drehen Sie den Knopf eine Position zur\"{u}ck.
\end{itemize}
Jetzt k\"{o}nnen Sie den Drucker an den Computer anschliessen.
Falls Sie den Drucker an einen Macintosh Computer anschliessen, fahren Sie mit den Instruktionen im Kapitel 3 fort. Falls Sie einen anderen Computer benutzen, fahren Sie fort mit Kapitel 4.
----Epoch: Etrans:SourceText 3:40pm Mail (Fundamental)----All----
EtransSourceView.dvi
Druckdichte Einstellung
Die gedruckte Seite sollte von exzellenter Qualität sein. Es gibt aber eine Reihe von Umweltfaktoren, wie hohe Temperatur und Feuchtigkeit, die Variationen in der Druckdichte verursachen können.
Falls die Testseite zu hell oder zu dunkel aussieht, verstellen Sie die Druckdichte am Einstellknopf an der linken Seite des Druckers (Figur 2-25).
Einstellung der Druckdichte:
Drehen Sie den Knopf ein oder zwei Positionen in Richtung des dunklen Indikators.
Schalten Sie den Drucker für einen Moment aus und dann wieder ein, so daß die Testseite gedruckt wird.
Wiederholen Sie die beiden vorherigen Schritte solange, bis Sie grau auf dem Blatthintergrund sehen, ähnlich wie bei leicht unsauberen Kopien eines Photokopierers.
Drehen Sie den Knopf eine Position zurück.
Jetzt können Sie den Drucker an den Computer anschliessen.
Falls Sie den Drucker an einen Macintosh Computer anschliessen, fahren Sie mit den Instruktionen im Kapitel 3 fort. Falls Sie einen anderen Computer benutzen, fahren Sie fort mit Kapitel 4.
EtransTermbase @ llsun6
Etrans:RawOutput @ llsun6
Etrans:SourceText @ llsun6
Etrans:RevisedOutput @ llsun6
xwordnet
Enter Search Word:
Sense Number:
density
Searches:
Noun
Verb
Adjective
Adverb
Options
Synonyms / Hypernyms of Noun density
2 senses of density
xterm
79 ~doug/Etrans>

During post-editing, the source text and target text can be displayed on alternate lines, which permits easy editing of the target text. This can be seen in the window at the top left of the screen. Below this are windows and icons for on-line dictionaries and termbanks, the source text alone, and the edited target text, etc. The window on the right shows the source text as it was originally printed.

Figure 2.1 Translators' Workstation while Post-Editing a Translation

Note that ETRANS has left the document format completely unaltered. It may be that the translation is actually slightly longer (or shorter) than the source text; any necessary adjustment to the pagination of the translation compared to the source is a matter for the document processing system.

After post-editing the remaining text, you have almost completed the entire translation process. Since it is not uncommon for translators to miss some small translation errors introduced by the MT system, you observe company policy by sending your post-edited electronic text to a colleague to have it double-checked. The result will be something like that on page 25.

Post-edited translation

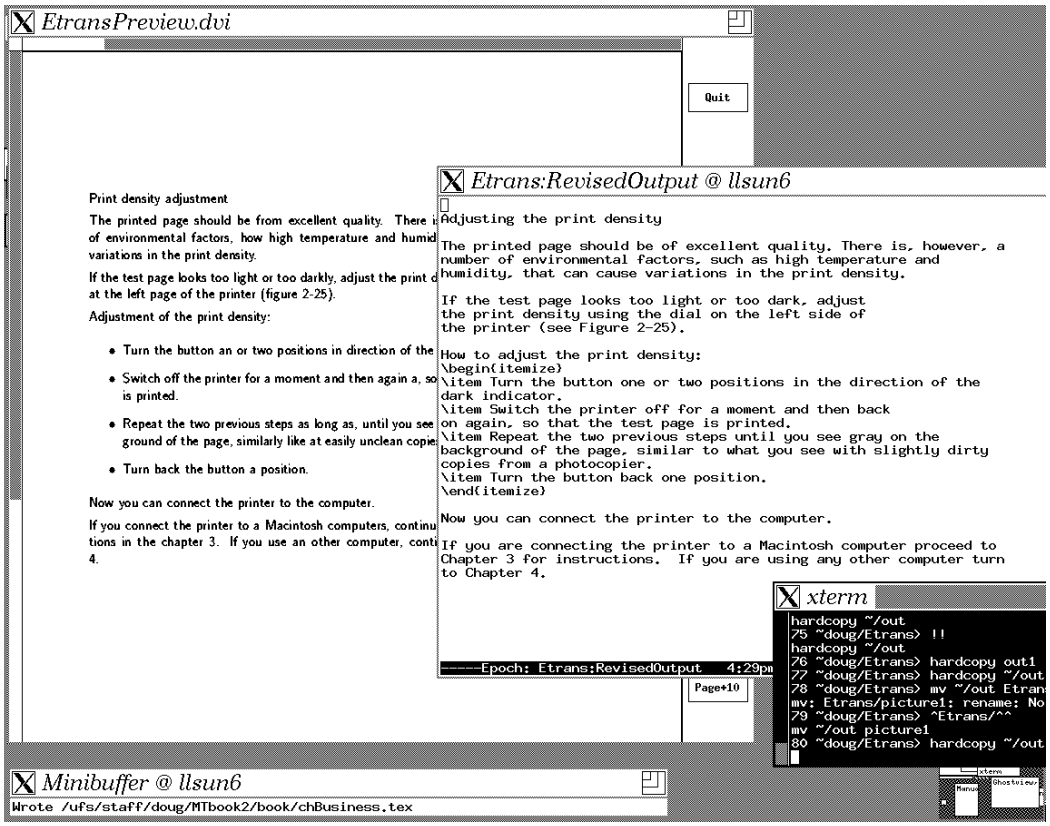
Adjusting the print density
The printed page should be of excellent quality. There is, however, a number of environmental factors, such as high temperature and humidity, that can cause variations in the print density.
If the test page looks too light or too dark, adjust the print density using the dial on the left side of the printer (see Figure 2-25).
How to adjust the print density:

- Turn the button one or two positions in the direction of the dark indicator.
- Switch the printer off for a moment and then back on again, so that the test page is printed.
- Repeat the two previous steps until you see gray on the background of the page, similar to what you see with slightly dirty copies from a photocopier.
- Turn the button back one position.

Now you can connect the printer to the computer.
If you are connecting the printer to a Macintosh computer proceed to Chapter 3 for instructions. If you are using any other computer turn to Chapter 4.

The only thing left to be done is to update the term dictionary, by adding any technical terms that have appeared in the document with their translation terms which other translators should in future translate in the same way, and report any new errors the MT system has committed (with a view to the system being improved in the future).

So that, in outline, is how MT fits into the commercial translation process. Let us review the individuals, entities and processes involved. Proceeding logically, we have as individuals:


EtransPreview.dvi
Quit
Print density adjustment
Adjustment of the print density:
Now you can connect the printer to the computer.
Page+10
Etrans:RevisedOutput @ llsun6
Adjusting the print density
The printed page should be of excellent quality. There is, however, a number of environmental factors, such as high temperature and humidity, that can cause variations in the print density.
If the test page looks too light or too dark, adjust the print density using the dial on the left side of the printer (see Figure 2-25).
How to adjust the print density:
\begin{itemize}
\item Turn the button one or two positions in the direction of the dark indicator.
\item Switch the printer off for a moment and then back on again, so that the test page is printed.
\item Repeat the two previous steps until you see gray on the background of the page, similar to what you see with slightly dirty copies from a photocopier.
\item Turn the button back one position.
\end{itemize}
Now you can connect the printer to the computer.
If you are connecting the printer to a Macintosh computer proceed to Chapter 3 for instructions. If you are using any other computer turn to Chapter 4.
-----Epoch: Etrans:RevisedOutput 4:29pm
xterm
hardcopy ~/out
75 ~doug/Etrans> !!
hardcopy ~/out
76 ~doug/Etrans> hardcopy out1
77 ~doug/Etrans> hardcopy ~/out
78 ~doug/Etrans> mv ~/out Etran
mv: Etrans/picture1: rename: No
79 ~doug/Etrans> ^Etrans/^^
mv ~/out picture1
80 ~doug/Etrans> hardcopy ~/out
Minibuffer @ llsun6
Wrote /ufs/staff/doug/MTbook2/book/chBusiness.tex

Having finished revising the translation, the result can be checked. One of the windows contains a preview of how the revised target text will look when it is printed. The other contains the revised translation, which can be edited for further corrections.

Figure 2.2 Translators' Workstation Previewing Output

- Documentation managers, who specify company policy on documentation.
- Authors of texts who (ideally) write with MT in mind, following certain established guidelines.
- Translators who manage the translation system in all respects pertaining to its day to day operation and its linguistic performance.

In many cases the document management role will be fulfilled by translators or technical authors. For obvious reasons, there will be fairly few individuals who are both technical authors and translators.

The important entities in the process are:

- Multi-Lingual Electronic Documents which contain text for translation.
- The Document Preparation system which helps to create, revise, distribute and archive electronic documents.
- The Translation System which operates on source text in a document to produce a translated text of that document.

Clearly any translation system is likely to be a very complex and sophisticated piece of software; its design at the linguistic level is discussed in detail in other chapters in this book. A detailed discussion of Electronic Documents can be found in Chapter 8.

Finally, the various processes or steps in the whole business are:

- Document Preparation (which includes authoring and pre-editing).
- The Translation Process, mediated by the translation system, perhaps in conjunction with the translator.
- Document Revision (which is principally a matter of post-editing by the translator).

The scenario gave a brief flavour of all three steps. We shall now examine each of them in rather more detail.

2.3 Document Preparation: Authoring and Pre-Editing

The corporate language policy as described in the scenario tries to ensure that text which is submitted to an MT system is written in a way which helps to achieve the best possible raw MT output. A human translator will often be able to turn a badly written text into a well written translation; an MT system certainly will not. Bad input means bad output. Exactly what constitutes *good* input will vary a little from system to system. However, it is

Basic Writing Rules

- Keep sentences short.
- Make sure sentences are grammatical.
- Avoid complicated grammatical constructions.
- Avoid (so far as possible) words which have several meanings.
- In technical documents, only use technical words and terms which
  - are well established, well defined and known to the system.

easy to identify some simple writing rules and strategies that can improve the performance of almost any general-purpose MT system. Here are some example rules:

Our example rules indicate sentences should be short. This is because MT systems find it difficult to analyse long sentences quickly or — more importantly — reliably. Lacking a human perspective, the system is always uncertain about the correct way to analyse a sentence; as the sentence gets longer, the number of uncertainties increases rather dramatically.

Sentences should also be grammatical, and at the same time not contain very complicated grammatical constructions. Whether or not an MT system uses explicit grammatical rules in order to parse the input, correct, uncomplicated sentences are always easier to translate

Some MT systems use linguistic knowledge to analyse the input sentences, others do not. In both cases correct, uncomplicated input sentences will enhance the translation performance because unnecessary translation problems are avoided. For example, the second piece of text below is more likely to be successfully translated than the first:

(3) New toner units are held level during installation and, since they do not as supplied contain toner, must be filled prior to installation from a toner cartridge.

(4) Fill the new toner unit with toner from a toner cartridge. Hold the new toner unit level while you put it in the printer.

The subclauses in the first sentence have been separated out as independent sentences in the second piece of text. The latter gives the instructions as a simple series of imperatives, ordered in the same way as the operations themselves.

The two final points in the list of writing rules prevent mistranslations by reducing potential sources of ambiguity. Many MT systems can do a reasonable job of selecting a correct interpretation of an ambiguous word in some circumstances, but they are unlikely to do this

successfully in all cases. (For example, ETRANS failed to get the correct interpretation of the two different occurrences of *Seite* (i.e. 'side' or 'page') in the passage above.) Problems of ambiguity are extensively discussed in later chapters.

Restricting MT input according to simple writing rules like the ones given above can greatly enhance the performance of an MT system. But this is not the only advantage: it can also improve the understandability of a text for human readers. This is a desirable feature in, for example, technical texts and instruction manuals. As a consequence, several large companies have developed and extended the idea of writing rules, including limited vocabulary, in order to produce restricted forms of English suitable for technical texts. These restricted forms are known as *controlled languages*. We will discuss controlled languages in detail in Chapter 8.

In the past few years special tools have become available for supporting the production of text according to certain writing rules. There are spelling checkers and grammar checkers which can highlight words that are spelled incorrectly, or grammatical errors. There are also **critiquing systems** which analyse the text produced by an author and indicate where it deviates from the norms of the language. For example, given the example above of an over-complex sentence in a printer manual, such a tool might produce the following output:

Text Critique

New toner units are held level during installation and, since they do not as supplied contain toner, must be filled prior to installation from a toner cartridge.

Sentence too long.
during installation — disallowed use of word: installation.
prior — disallowed word.
since — disallowed clause in middle of sentence.

This is a rather sophisticated analysis of various violations found in the sentence. The controlled language this critiquing system is designed for only sanctions the word *installation* if it refers to some concrete object, as in *Remove the forward wheel hydraulic installation*; in this particular case *installation* is being used to denote the process of installing something. For the time being, this type of analysis is too advanced for most critiquing systems, which would find the sentence too difficult to analyse and would simply note that it is too long, not analysable, and contains the unknown word *prior*.

Critiquing systems ensure that texts are written according to a set of writing rules or the rules of a controlled language and thus help to catch errors which might upset an MT

system. As a consequence they reduce the amount of time necessary for post-editing machine translated texts. They also reduce the time that someone else would normally have to spend on checking and revising the input text.

There is no theoretical reason why a controlled language critiquing system could not be completely integrated with an MT system designed to handle the controlled language — so that the translation system itself produces the critique while analysing the text for the purpose of translation. In fact, if the MT system and the critiquing system are completely separate, then the same piece of text will always have to be analysed twice — once by the critiquing system and a second time by the MT system. Moreover, the separation means that the same controlled language rules and electronic dictionary entries are repeated twice — once for each component. This makes it more expensive to revise or alter the controlled language. For these reasons, we can expect that MT system suppliers will seek to integrate controlled language critiquing and controlled language MT as closely as possible.

Of course, in practice not all text submitted to MT systems is (or can be, or should be) written according to a set of writing rules. Although this is not necessarily problematic it should be borne in mind that the less a text conforms to the rules mentioned above, the worse the raw translation output is likely to be. There will be a cutoff point where the input text is so badly written or so complicated that the raw output requires an uneconomically large amount of post-editing effort. In this case it may be possible to rewrite the problematic sentences in the input text or it may prove simplest to do the whole thing by hand.

2.4 The Translation Process

In the scenario we sketched above, the source text or some selected portion thereof was passed to the translation system which then produced raw translated output without any further human intervention. In fact, this is merely one of many ways the translation step can proceed.

2.4.1 Dictionary-Based Translation Support Tools

One point to bear in mind is that translation support can be given without actually providing full automatic translation. All MT systems are linked to electronic dictionaries which, for the present discussion, we can regard as sophisticated variants of their paper cousins. Such electronic dictionaries can be of immense help even if they are supplied or used without automatic translation of text. Here is one possible scenario:

> You are translating a text by hand. Using a mouse or the keyboard, you click on a word in the source text and a list of its possible translations is shown on screen. You click on the possible translation which seems most appropriate in the context and it is inserted directly into the target language text. Since you usually do this before you start typing in the translation of the sentence which contains the unknown work, the

> inserted word is inserted in the middle of an otherwise blank target language sentence. You then type in the rest of the translation around this inserted word.

Since technical texts typically contain contain large number of terms, and their preferred translations are not always remembered by the translator, this simple form of support can save a lot of time. It also helps to ensure that terms are consistently translated.

This *click to see, click to insert* facility is useful in dealing with low-frequency words in the source text. In technical text, technical terms — which can be complex multi-word units such as *faceplate delivery hose clip* — will usually have only one translation in the target language. If the electronic dictionary has a list of terms and their translations, those translations can be directly inserted into the target text. This gives the following scenario:

> You are translating a technical text by hand. You click on the icon *Term Support* and all the source language terms in the current text unit which are recognised as being in the electronic term dictionary are highlighted. A second click causes all the translations of those terms to be inserted in otherwise empty target language sentences. You then type in the rest of the translation around each inserted term.

Translation Aids in the Workplace No. 72:
Automatic Lexical Lookup

Dictionary-based translation support tools of this sort depend on two things:

1 The required terms and words must be available in the electronic dictionary. This

may well require that they were put there in the first place by translators in the organization using the tool.

2 There must be some simple means for dealing with the inflections on the ends of words since the form of a word or term in the text may not be the same as the cited form in the dictionary. As a simple example, the text may contain the plural form *faceplate delivery hose clips* rather than the singular form kept in the dictionary. The problem is more complex with verb inflections and in languages other than English.

These and other issues concerning the MT dictionary will be discussed in Chapter 5.

2.4.2 Interaction in Translation

MT systems analyse text and must decide what its structure is. In most MT systems, where there are doubts and uncertainties about the structure, or about the correct choice of word for a translation, they are resolved by appeal to in-built *rules-of-thumb* — which may well be wrong for a particular case. It has often been suggested that MT systems could usefully interact with translators by pausing from time to time to ask simple questions about translation problems.

Another sort of interaction could occur when the system has problems in choosing a correct source language analysis; a good analysis is needed to ensure good translation. For example, suppose that a printer manual being translated from English contains the following sentence:

(5) Attach the printer to the PC with a parallel interface cable.

The question is: are we talking about a particular type of PC (personal computer) which comes with a *parallel interface cable* (whatever that is) or any old PC which can be connected to the printer by means of an independent *parallel interface cable*? In the first case, the *with*, in the phrase *with a parallel interface cable* means *having* or *fitted with* and modifies the noun *PC*, whilst in the second it means *using* and modifies the verb *attach*. One good reason for worrying about the choice is because in many languages *with* will be translated differently for the two cases. Faced with such an example, an MT system might ask on screen exactly the same question:

(6) Does *with a parallel interface cable* modify *the PC* or does it modify *Attach*?

Another sort of analysis question arises with pronouns. Consider translating the following:

(7) Place the paper in the paper tray and replace the cover. Ensure that it is completely closed.

Does *it* in the second sentence refer to *the paper*, *the paper tray*, or *the cover*? The decision matters because the translation of *it* in many languages will vary depending on the gender of the expression it refers back to. Making such a decision depends on rather

subtle knowledge, such as the fact that covers, but not trays or paper are typical things to be closed, which is hard perhaps impossible to build into an MT system. However, it is the sort of question that a human translator may be able to answer.

The following is a possible scenario:

> You are translating a text interactively with an MT system. The system displays the source text in one window, while displaying the target text as it is produced in another. On encountering the word *it*, the system parses, highlights the words *paper*, *paper tray*, and *cover* in the first sentence, and asks you to click on the one which is the antecedent (i.e. the one *it* refers back to). It is then able to choose the appropriate form of the translation, and it proceeds with the rest of the sentence.

It is hardly surprising that a machine may need to ask such questions because the answers may not be at all clear, in some cases even for a human translator. With poorly written technical texts, it may even be the case that only the author knows.

2.5 Document Revision

The main factor which decides the amount of post-editing that needs to be done on a translation produced by machine is of course the quality of the output. But this itself depends on the requirements of the client, in particular (a) the translation aim and (b) the time available. In the case of the printer manual in the scenario above the translation aim was to provide a printer manual in English for export purposes. The fact that the translation was going to be widely distributed outside the organization required it to be of high quality — a correct, well-written and clear piece of English text, which means thorough and conscientious post-editing.

The opposite situation occurs when a rough and ready translation is needed out of some language for personal or internal use, perhaps only to get the gist of some incoming text to see if it looks interesting enough for proper translation. (If it is not, little time or money or effort has been wasted finding out). Here is the sort of scenario in which it might work:

> You are an English-speaking agronomist monitoring a stream of information on cereal crop diseases coming in over global computer networks in four different languages. You have a fast MT system which is hooked into the network and translates — extremely badly — from three of the languages into English. Looking at the output and using your experience of the sort of things that reports contain, you should be able to get enough of an idea to know whether to ignore it or pass it on to your specialist translators.

Of course, in this situation it is the speed of the MT system, not its quality that matters —

a very simple system that does no more than transliterate and translate a few of the words may even be enough.

We've now looked at two cases: one in which full post-editing needed to be done, one in which no post-editing whatsoever was required. Another option could be to do some post-editing on a translation in order to make it easy to read and understand, but without having the perfection of a published text in mind. Most post-editors are also translators and are used to producing high quality texts. They are likely to apply the same sort of output standards to their translations produced automatically. Though this policy is very desirable for, for instance, business correspondence and manuals, it is not at all necessary to reach the same sort of standard for internal electronic mail. Some MT output could be subject to a *rough and ready* post-edit — where the post-editor tries to remove or adjust only the grossest errors and incomprehensibilities — rather than the usual thorough and painstaking job. The main advantage of this option is that translator time is saved. Even if documents are occasionally sent back for re-translation or re-editing, the *rough and ready* post-edit policy might still save money overall. Again, the factors of translation aim and time available play an important role.

MT systems make the same sorts of translation mistake time and time again. Sometimes these errors can be eliminated by modifying the information in the dictionary. Other sorts of errors may stem from subtle problems in the system's grammars or linguistic processing strategies which cannot ordinarily be resolved without specialist knowledge. Once an error pattern has been recognised, a translator can scan text looking for just such errors. If the error is just a matter of consistently mistranslating one word or string of words, then — as in the scenario — the ordinary search-and-replace tools familiar from word processors will be of some help. In general, since the errors one will find in machine translated texts are different from those one finds in other texts, specialized word processor commands may be helpful. For example, commands which transpose words, or at a more sophisticated level, ones which change the form of a single word, or all the words in a certain region from masculine to feminine, or singular to plural, might be useful post-editing tools.

The imaginary company that we have been discussing in the previous sections deals with large volumes of similar, technical text. This text similarity allows the MT system to be tuned in various ways, so as to achieve the best possible performance on one particular type of text on one particular topic. An illustration of this can be found in the section heading of our example text *Einstellung der Druckdichte*. The German word *Einstellung* can have several translations: *employment, discontinuation, adjustment* and *attitude*. Since we are dealing here with technical texts we can discard the first and last possible translations. Of the two translations left, *adjustment*, is the most common one in this text type, and the computer dictionaries as originally supplied have been updated accordingly. The tuning of a system takes time and effort, but will in the long run save post-editing time.

Obviously enough, the difficulty of post-editing and the time required for it correlates with the quality of the raw MT output: the worse the output, the greater the post-edit effort. For one thing, the post-editor will need to refer more and more to the source language text when the output gets less intelligible. Even though this seems to be a major drawback

at the beginning, bear in mind that post-editors will get used to the typical error patterns of the MT system; MT output that may seem unintelligible at the beginning will require less reference to the source language text after some time. Familiarity with the pattern of errors produced by a particular MT system is thus an important factor in reducing post-editing time. More generally, familiarity with the document processing environment used for post-editing and its particular facilities is an important time saver.

2.6 Summary

This chapter has given a picture of how MT might be used in an imaginary company, and looked in outline at the typical stages of translation: *document preparation*, *translation* (including various kinds of human involvement and *interaction*), and *document revision*, and at the various skills and tools required. In doing this we have tried also to give an idea of some of the different situations in which MT can be useful. In particular, the case of 'gist' translation, where speed is important, and quality less important, compared to the case where a translation is intended for widespread publication, and the quality of the finished (post-edited) product is paramount. These are all matters we will return to in the following chapters.

2.7 Further Reading

Descriptions of how MT is actually used in corporate settings can be found in the Proceedings of the Aslib Conferences (normally subtitled *Translating and the Computer*) which we mentioned in the Further Reading section of Chapter 1.

For readers interested in finding out more about the practicalities of pre- and post-editing , there are several relevant contribution in Vasconcellos (1988), in Lawson (1982a). There is a useful discussion of issues in pre-editing and text preparation, in Pym (1990), and we will say more about some related issues in Chapter 8.

An issue that we have not addressed specifically in this chapter is that of machine aids to (human) translation, such as on-line and automatic dictionaries and terminological databases, multilingual word processors, and so on. We will say more about terminological databases in Chapter 5. Relevant discussion of interaction between machine (and machine aided) translation systems and human users can be found in Vasconcellos (1988),Stoll (1988),Knowles (1990) and various papers by Alan Melby, including Melby (1987, 1992), who discusses the idea of a 'translator's workbench'. In fact, it should be clear that there is no really hard and fast line that can be drawn between such things and the sort of MT system we have described here. For one thing, an adequate MT system should clearly include such aids in addition to anything else. In any case, in the kind of setting we have described, there is a sense in which even an MT system which produces very high quality output is really serving as a translators' aid, since it is helping improve their productivity by producing draft translations. What are sometimes called distinction between 'Machine Aided Human Translation', 'Human Aided Machine Translation', and 'Machine Translation' *per se* actually form a continuum.